联合国可持续发展目标出版物检索策略研究 *

于　曦
（天津师范大学图书馆，天津　300387）

摘　要：［**目的 / 意义**］联合国可持续发展目标（SDGs）出版物，是衡量国家、地区、机构对可持续发展目标相关研究贡献程度的文献基础支撑，研究优化出版物检索策略可为开展相关工作提供方法参考。［**方法 / 过程**］基于 Elsevier 和 Clarivate 检索系统平台，以可持续发展目标相关论文作为检索对象，比较分析两个数据库检索结果的差异以及产生差异的原因，并针对两种查询方法的优势和不足提出改进建议。［**结果 / 结论**］相较于传统的学科与文献分类体系，联合国可持续发展目标出版物分类是一种未经作者和编辑参与的分类体系，目前很难建立出版物与检索文献之间的精准映射关系，因此需要充分认知不同检索系统平台的分类差异，优化可持续发展目标出版物检索策略，充分挖掘与利用图书馆的数字资源，为教学科研提供面向可持续发展目标的文献资源支撑。

关键词：可持续发展目标（SDGs）出版物　检索策略　Elsevier　Clarivate
分类号：G252.6
DOI：10.31193/SSAP.J.ISSN.2096-6695.2025.01.11

0 引 言

2015 年 9 月联合国通过了《2030 年可持续发展议程》，首次提出包含 17 项内容的“可持续发展目标”（Sustainable Development Goals, SDGs）。联合国可持续发展目标（SDGs）是一项全球行动倡议，旨在消除贫困、保护环境、确保繁荣与和平，为解决社会、经济和环境中最关键的问题提供了一个全球框架与指标体系，将可持续发展目标分解为更多的颗粒级别，并以此为基准衡量可持续发展目标的全球进展[1]。

联合国可持续发展目标的提出，引起了世界各国政府组织和研究机构的重视，它们都希望能为解决全球问题作出自己的贡献。美国国际开发署（U.S. Agency for International Development, USAID）致力于推动实现可持续发展目标的进程[2]。澳大利亚在《2017 年外交政策白皮书》中强调，澳大利亚有责任为全球减贫、减轻痛苦和促进可持续发展努力作出贡献。澳大利亚政府

* 本文系天津市哲学社会科学规划（重点）项目“‘双一流’学科建设背景下全域学科分类映射理论与实践”（项目编号：TJTQ21-001）的研究成果之一。
［作者简介］于曦，女，副研究馆员，研究方向为学科服务、数据分析，Email：yuxisd_2004@126.com。

联合澳大利亚的大学、企业和民间组织，共同应对《2030 年可持续发展议程》提出的风险和机遇[3]。我国将联合国可持续发展目标纳入国家“十三五”发展规划，发布《中国落实 2030 年可持续发展议程国别方案》[4]，建设可持续发展大数据国际研究中心[5]。在高校可持续发展目标人才培养层面，奥克兰大学（University of Auckland）针对可持续发展目标开设了一系列本科生和研究生教学课程[6]。浙江大学为解决可持续发展目标下的环境治理问题开设了全球暑期学校课程[7]，复旦大学根据联合国提出的 17 项可持续发展目标对学校的教学课程、科研成果和实践活动进行分类，制定了可持续发展行动报告[8]。这些高校开设的可持续发展目标相关课程涉及学科主题广泛，而基于传统学科分类法的文献组织和检索策略基本上无法获取相关的研究资源。

需要说明的是，本文所指科学出版物属于狭义概念，仅限于期刊出版物中的学术论文。将科学出版物有效映射到相应的联合国可持续发展目标，是开展可持续发展目标相关研究的文献基础。如何将科学期刊出版物与联合国可持续发展目标的大类对应起来，从而检索获取可持续发展目标相关出版物则显得至关重要。为落实联合国可持续发展目标，各国政府纷纷成立可持续发展研究机构，高校也根据不同的可持续发展目标设立可持续发展目标相关教学课程，开展有关可持续发展目标的学术研究。高校图书馆作为高校教学科研的辅助机构，应提供针对可持续发展目标相关课程和学术研究的文献资源支持服务。例如，高校图书馆可以分析学校的教学内容和科研方向主要针对哪几个可持续发展目标，调查与这些可持续发展目标相关的文献资源覆盖率，开展基于可持续发展目标的馆藏资源建设和评估。但可持续发展目标分类不同于以往任何一种分类体系，也不依赖于现有任何一种分类方法。如何建立文献资源与可持续发展目标的对应关系并精确检索出版物，是图书馆开展上述服务工作的前提和基础。

本文通过梳理可持续发展目标出版物检索方法与获取现状，在 Elsevier 和 Clarivate 两个检索平台中分别检索某高校的可持续发展目标相关出版物，比较两个检索平台的检索结果，找出产生差异的原因。针对两个检索平台对可持续发展目标出版物检索的方法缺陷，提出一些可行的改进方案，旨在为可持续发展目标出版物查询研究提供参考。

1 联合国可持续发展目标出版物相关检索方法

自从联合国提出可持续发展目标，一些学者、信息服务机构和学术研究机构就开展了相关出版物的识别和检索研究。目前，围绕联合国可持续发展目标相关主题出版物的常用检索方法主要包括以下三种。

第一种是布尔逻辑查询方法。开展布尔逻辑查询的前提是生成相关主题词，即本体构建。然后对主题词进行单独或组合使用，再将布尔逻辑查询条件应用于出版物的标题、摘要、关键词等文本中。Armitage 等[9]应用布尔逻辑方法，将检索对象限制在对联合国可持续发展目标有直接贡献的出版物，并尽量减少所提问题对布尔逻辑的影响，从而产生针对性更强的检索结果。Bordignon[10]将检索范围限定在全部科学期刊分类，以此来缩小关键词的多义现象，并使用

CorTexT 文本挖掘工具来充实选定的出版物。奥克兰大学使用文本挖掘技术和 n-gram 分析方法从出版物元数据中提取全局和局部关键词，根据包含关键词的出版物数量和“关键词频率—逆文档频率（TF-IDF）”的值进行排序[11]，又根据本地研究人员提供的信息查询，创建了一个本地化版本，考虑了更多针对澳大利亚和新西兰研究主题的论文[12]。澳大利亚、新西兰和太平洋地区的可持续发展解决方案网络（Sustainable Development Solutions Network, SDSN）为可持续发展目标映射提供了一份“可持续发展目标关键词汇编列表”[13]，该关键词列表由澳大利亚、新西兰和太平洋地区的几所大学联合开发，其中莫纳什大学（Monash University）开发了最初的关键词列表，根据 SDSN 和同行大学的输入对列表进行了修改，总共产生 915 个可持续发展目标特定关键词，涵盖所有 17 个可持续发展目标。奥地利研究人员使用来自奥地利 13 所大学的内部研究文献系统的数据集，将奥地利研究出版物映射到可持续发展目标中[14]。Confraria 等[15]从政策报告、出版物、论坛等获得可持续发展目标特定术语，并对术语进行选择处理，使用这些术语来识别基于引用的出版物集。

第二种方法是通过机器学习，采用监督方式（分类）或无监督方式（聚类）将相关出版物映射到联合国可持续发展目标。监督方法通常采用相同的出版物查询来获得标记数据集作为训练模型[16-17]。无监督方法通常使用出版物文本表示或引用图谱来完成，产生的聚类随后直接或经由中间聚类（主题）映射到相应的可持续发展目标[18]。联合国使用主题建模算法，采用无监督模式为经济和社会事务部的出版物建立可持续发展目标分类器[19]。Hajikhani 等[20]利用科学与技术之间的相互关系，通过联合国可持续发展目标出版物分类编译机器学习模型，对可持续发展目标相关专利文献进行分类，揭示可持续发展目标在专利中的表现情况。Zhang 等[21]将 4 个映射研究应用于深度学习方法中绘制可持续发展目标映射。Digital Science 根据联合国对可持续发展目标的定义提取关键术语，定义每个目标的关键字搜索字符串，其旗下的 Dimensions 平台为每个可持续发展目标都创建了一个训练集。Dimensions 平台应用自然语言处理和机器学习的方法，对 17 个可持续发展目标训练集产生分类方案[22]。Digital Science 与 Google Cloud 合作，将 Dimensions 与 Google Cloud 的 BigQuery 集成，使得 Dimensions 数据库可以进行大规模分析，采用 Dimensions 可持续发展目标分类器对英国“卓越研究框架”（Research Excellence Framework, REF）进行案例分析[23]。

第三种方法是将布尔逻辑查询和机器学习相结合，识别与获取相关主题出版物。欧洲大学联盟使用了关键词组合、布尔和邻近运算符等方法，研发出 169 个目标级别的可持续发展目标查询方法[24-25]，构建出将多语种学术论文映射到联合国出版物的机器学习模型[26]。Elsevier 联合南丹麦大学（University of Southern Denmark）、欧洲大学联盟和奥克兰大学，研究出将科学出版物映射到可持续发展目标的关键词查询方法，并辅助机器学习大大提升查找的召回率和准确率，该方法被《泰晤士高等教育》（Time Higher Education，THE）推出的大学影响力排名所采用[27]。

联合国可持续发展目标特定出版物的检索方法，主要包括“关键词查询”、“机器学习模型”及“关键词查询 + 机器学习模型”三种类别（表 1）。

表 1　联合国可持续发展目标出版物查询方法 [28]

类别	获取方法	方法描述
关键词查询	奥克兰查询 v2	为更好地了解研究贡献，奥克兰大学的可持续发展目标关键词词典项目在联合国和 THE 开发的流程基础上，创建可用于识别可持续发展目标相关出版物的扩展关键词列表
	Aurora 查询 v5	使用关键词组合、布尔和邻近运算符等方法，开发可持续发展目标查询方法，将研究出版物映射到可持续发展目标
	卑尔根可持续发展目标查询	构建部分可持续发展目标的出版物查询方法，该查询方法已在 Scopus 中运用，样本数据应用结果表现较好；样本中还出现其他出版物，这些出版物作为反面例子没有被查询到；创建两个数据集，一个基于手动方法查询，另一个基于主题方法查询，分别称为“卑尔根 TAA”和“卑尔根 TBA”
	Elsevier 查询 2020	确定了支持联合国可持续发展目标的关键词
	SIRIS 可持续发展目标查询	从联合国官方的可持续发展目标、具体指标清单以及围绕可持续发展目标相关出版物中提取关键词术语；查询系统通过预训练的词向量模型和从维基百科中选择相关词语的算法进行扩展；每个可持续发展目标有多个查询
	Bordignon 可持续发展目标查询	使用所有科学的期刊分类体系，限制来自 Elsevier 2020 查询中的关键词所属主题领域，从而减少术语的多义性
机器学习模型	Aurora ML v0.2	通过人工智能将多语言学术论文映射到联合国可持续发展目标
	南非可持续发展目标中心	构建将文本映射到可持续发展目标的机器学习模型
	Digital Science 可持续发展目标	Digital Science 使用 Dimensions 针对每个可持续发展目标都生成一个训练集，应用自然语言处理和机器学习产生分类方案
关键词查询 + 机器学习模型	Elsevier 查询 +ML 2021	改进 Scopus 和 Aurora 查询，支持 2021 年联合国可持续发展目标的研究
	Elsevier 查询 +ML 2022	Elsevier 查询 +ML 2021 的简化版本，在 SDG3 中增编“新型冠状病毒”的相关内容
	Elsevier 查询 +ML 2023	使用与 Elsevier 2022 可持续发展目标映射完全相同的搜索查询和机器学习算法，对五个可持续发展目标进行微修改；通过删除基于期刊标识符的排除列表来缩短查询时间，这些排除列表用以过滤掉不属于可持续发展目标的期刊论文

2　Elsevier 与 Clarivate 检索平台案例分析

根据联合国可持续发展目标出版物检索方法研究现状，本文基于 Elsevier 与 Clarivate 两个不同检索平台，对获取可持续发展目标相关出版物的检索方法进行案例分析与比较研究。

2.1　检索平台及案例选取

本文选取 Elsevier 与 Clarivate 作为联合国可持续发展目标出版物检索分析平台。Elsevier 是全球最大的出版集团，其旗下的 Scopus 和 Scival 致力于为学者、科研机构和管理部门提供文献检索、科学计量分析与评价服务。Clarivate 旗下的 Incites 提供科学计量指标和引文分析服务。Elsevier 与 Clarivate 两个检索平台收录文献范围比较全面，质量相对较高，统计分析功能强大，可以作为科学研究的资源保障以及学科数据、绩效评价的分析工具。Elsevier 和 Clarivate 都研发出联合国可持续发展目标与科学出版物之间的映射方法，包括从第 1 大类到第 16 大类的可持续发展目标映射方法，但均未开展第 17 大类的映射研究。因此，开展 Elsevier 和 Clarivate 对可持续发展目标科学出版物识别的对比研究具有实际意义。

本文选取某高校发表的相关出版物作为案例分析对象，该高校属于综合类“双一流”本科院校，涉及理、工、经、管、文、法、医、教育、艺术等多个学科，学科覆盖范围广且科研成果数量多，能满足研究样本分析要求。以该高校 2018—2022 年在 Elsevier 和 Clarivate 两个数据库中均收录的可持续发展目标相关论文为对象进行检索，检索时间为 2023 年 12 月 4 日，文献类型为研究论文和综述论文，检索平台分别为 Elsevier 的 Scival 平台和 Clarivate 的 Incites 平台。

2.2　在 Scival 和 Incites 平台的检索结果比较

检索结果显示，该高校 2018—2022 年在 Scival 和 Incites 两个数据平台中均被收录的可持续发展目标论文共有 2696 篇。联合国可持续发展目标（SDGs）共有 17 项内容大类（包括 SDG1、SDG2、SDG3、SDG4、SDG5、SDG6、SDG7、SDG8、SDG9、SDG10、SDG11、SDG12、SDG13、SDG14、SDG15、SDG16、SDG17），但这些论文内容在两个数据平台中归属的 SDGs 大类并不相同。不同大类在两个数据库中的分布存在交叉关系，几乎每一个大类在两个数据平台的分布都不完全重合（图 1）。

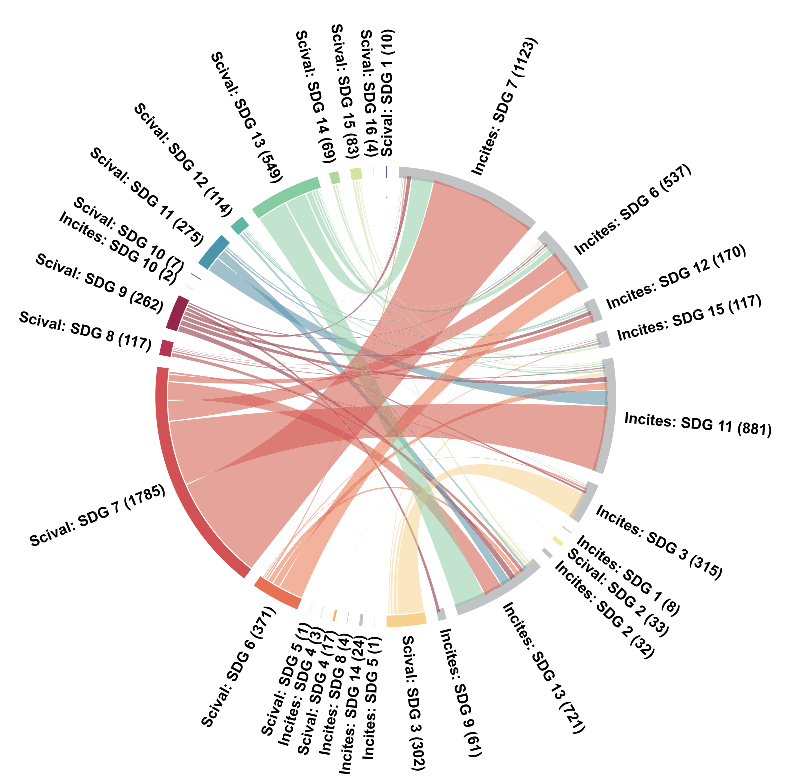


图 1　SDGs 论文在 Scival 和 Incites 平台的交叉分布

在统计每篇论文数时，有些涉及多个可持续发展目标的论文可能会被重复计算，因此相关论文总数会超过论文实际数量。在 2696 篇论文里，SDG7 在两个数据库中的论文数量均是最多的，说明该高校发表的可持续发展目标论文主要集中在 SDG7 领域，即“可负担的清洁能源”。鉴于此，下面以该高校发表的 SDG7 论文为研究对象，进一步分析两个数据平台检索结果的差

异。同时，统计出 SDG7 论文在 Scival 和 Incites 两个平台中的共有主题词数量与交叉分布情况（图 2）。

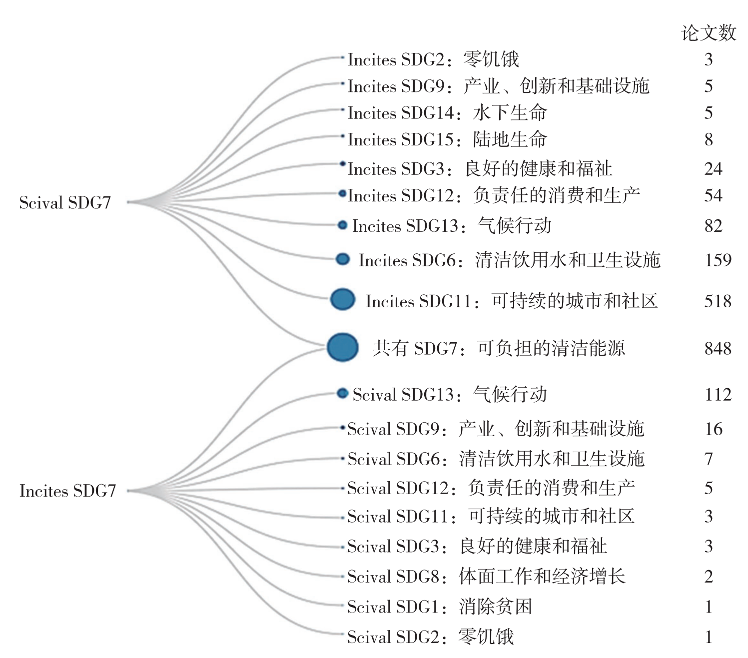


图 2　SDG7 论文在 Scival 和 Incites 两个平台中的数量分布

2.2.1　Scival 和 Incites 共有的 SDG7 论文主题词分析

（1）两平台共有 SDG7 论文主题词

实际上，在 Scival 和 Incites 中均被分类为 SDG7 的论文并不完全重合。据检索统计结果，在两库中均被划分为 SDG7 的论文共有 848 篇（图 2），占 Scival SDG7 论文总数的 51.02%，占 Incites SDG7 论文总数的 85.83%。这些共有的 SDG7 论文，其研究内容可能与两库中均包含的研究主题有关。

从 Scival 查找 SDG7 相关论文的主要主题词，并在 Incites 查找 SDG7 相关论文的微观主题。检索结果显示，Scival 平台的 SDG7 相关论文有 1000 个主要主题词，而 Incites 平台的 SDG7 相关论文仅对应 77 个微观主题。通过对比 Scival SDG7 主要主题词和 Incites SDG7 微观主题，发现两者共有 12 个实质上完全相同的主题词（表 2）。通过逐篇研读 848 篇共有 SDG7 论文的题名、摘要和关键词，发现多篇论文的研究内容与两平台中的共有主题词有关。

表 2　Scival 和 Incites 共有 SDG7 主题词

序号	Scival 论文主题词	Incites 论文主题词
1	doubly fed induction generator	doubly fed induction generator
2	ground source heat pump	ground source heat pump
3	solar air heater	solar air heater
4	bio–energy	bioenergy
5	dye sensitized solar cells	dye–sensitized solar cells

续表

序号	Scival 论文主题词	Incites 论文主题词
6	hydrogen storage	hydrogen storage
7	inertial confinement fusion (icf)	inertial confinement fusion
8	organic solar cells (oscs)	organic solar cells
9	perovskite solar cells (pscs)	perovskite solar cells
10	renewable energy (re)	renewable energy
11	solid oxide fuel cell (sofc)	solid oxide fuel cell
12	mppt	the maximum power point tracking
		the maximum power point tracking (mppt)

（2）两平台 SDG7 论文主题词差异

Scival SDG7 的主要主题词数量较多，但其中含有大量书写形式不同但意思完全一样的词语，如“dye sensitized solar cell（染料敏化太阳能电池）”（表 3）。

表 3　同一词语在 Scival 平台的不同书写形式

词语	不同书写形式
dye sensitized solar cell（染料敏化太阳能电池）	dye sensitized solar cell
	dye sensitized solar cells
	dye sensitized solar cells (dsscs)
	dye–sensitized solar cell (dssc)
	dye–sensitized solar cells (dscs)
	dye–sensitized solar cells (dssc)
	dye–sensitized solar cells (dsscs)
	dye–sensitized solar cells

这是因为 Scival 的主题词是通过文本挖掘技术从文献中提取的，文献中同一词语的不同书写形式都可能会被挖掘出来，为保证检索的全面性，主题词的所有书写形式都将被编辑到检索式中，而这种方法仅仅对已有文献生效，不能囊括新发表文献出现的书写形式。相比之下，Incites SDG7 的微观主题仅由一个词语来体现，该词语包含的内容则较为宽泛，书写形式也比较规范。

将 Scival SDG7 的主要主题词去除复数、缩写词、词之间的“–”和词尾的“.”，最后仅剩 609 个主题词。这 609 条记录中还存在大量针对同一个词语的多种应用形式，这是造成 Scival SDG7 有如此多的主题词的另一主要原因。排除掉不同书写形式和应用形式带来的差异，Scival SDG7 和 Incites SDG7 的主题词形式仍然存在很多差异。这里以“renewable energy”（可再生资源）为例，列出其在 Scival 平台的多种应用形式（表 4）。

2.2.2　Scival 自带的 SDG7 论文分析

如上文所述，Scival SDG7 论文中有 814 篇论文在 Incites 中未被划归到 SDG7，而是被划分到其他 SDGs 大类中，其中有 518 篇被划分到 SDG11（图 2）。推测 Scival SDG7 与 Incites SDG11

表 4　同一词语在 Scival 平台的多种应用形式

原始词	多种应用形式	
renewable energy（可再生资源）	renewable energy application	renewable energy power
	renewable energy applications	renewable energy production
	renewable energy consumption	renewable energy resources
	renewable energy conversion	renewable energy sources (res)
	renewable energy development	renewable energy sources (ress)
	renewable energy generation	renewable energy storage
	renewable energy integration	renewable energy supply
	renewable energy laboratory	renewable energy technology
	renewable energy penetration	renewable energy utilization

具有共同的主题词。查找 Scival SDG7 的论文主题词和 Incites SDG11 论文对应的微观主题，发现 Scival SDG7 中的主题词“lithium-sulfur batteries”（锂硫电池）和“lithium ion batteries（libs）”（锂离子电池）与 Incites SDG11 中的“2.62.616 Lithium-Sulfur Batteries”（锂硫电池）和“2.62.138 Lithium-Ion Battery”（锂离子电池）高度相似。Scival SDG7 的主题词“electric bus”（电动公交车）、“electric buses”（电动公交车）、“hybrid electric vehicle（hev）”（混合动力电动汽车）等与 Incites SDG11 中的“4.18.788 Electric Vehicles”（电动汽车）内容近似。此外，一些论文中还包括“electrode”（电极）、“anode”（阳极）、“cathode”（阴极）、“battery”或“batteries”（电池）等与“锂硫电池”和“锂离子电池”相关的词语，逐一研读这些论文内容，发现大部分论文与“锂硫电池”或“锂离子电池”相关。此外，还有一少部分是关于其他类型的电池，如锌离子电池和钠离子电池等，这些论文所属微观主题均为“2.62.52 Supercapacitor”（超级电容器），这是 Incites SDG11 包含的微观主题。在论文题名、摘要和关键词中查找上述提及的主题词，发现有 456 篇论文含有这些词语，仅有 62 篇论文没有。

对于未查到的 62 篇论文，通过逐篇筛查论文题名、关键词和摘要，发现这些论文大多含有“solar power”（太阳能）、“wind energy”（风能）、“geothermal system”（地热系统）等 Scival SDG7 论文的主题词，同样还含有“aerosols”（气溶胶）、“distributed storage systems”（分布式存储系统）、“eddy current testing”（涡流检测）、“internet of things”（物联网）、“supercapacitor”（超级电容器）等 Incites SDG11 论文包含的微观主题。

2.2.3　Incites 自带的 SDG7 论文分析

Incites SDG7 论文中有 140 篇论文在 Scival 中未被划归到 SDG7，而是被划分到其他 SDG 大类中，其中有 112 篇被划分到 SDG13（图 2）。据此推测，Incites SDG7 论文与 Scival SDG13 论文具有共同的主题词。查找 Incites SDG7 论文包含的微观主题和 Scival SDG13 论文包含的主要主题词并进行比较，发现 Incites SDG7 论文中包含的“2.62.2102 CO_2 reduction”（CO_2 减排）与 Scival SDG13 论文中的“CO_2 reduction”（CO_2 减排）完全相同。在 112 篇属于 Scival SDG13 的论文中，在其标题、关键词和摘要中大多含有“CO_2 reduction”（CO_2 减排）或“carbon dioxide reduction”（CO_2 减排）。在其他论文中，一般包括“greenhouse gas emission”（温室气体排放）、“global warming”（全球变暖）或“CO_2 electroreduction”（CO_2 电还原）等词语，这些词

语在 Scival SDG13 论文的主题词中均能被找到，可知这些论文在 Scival 中可能被划分到 SDG13 类别。

3 结果讨论与原因分析

排除 Elsevier 与 Clarivate 这两个数据库收录文献范围的不同，分别在其下属的 Scival 与 Incites 平台中检索关于联合国可持续发展目标的论文出版物，发现两者的检索结果存在较大差异。产生这些差异的具体原因大致有如下几点。

（1）分类存在不同的判断规则

Elsevier 对联合国可持续发展目标出版物的获取，主要通过构建相关论文的主题词及相应检索式来实现。早在 2015 年，联合国刚通过《2030 年可持续发展议程》的时候，Elsevier 就构建出最初版本的可持续发展目标主题词列表。此后不断改进查询策略，更新可持续发展目标主题词列表，完善可持续发展目标查询方法，通过查询将检索到的论文对应到相应的可持续发展目标大类，对检索到的论文根据所属的可持续发展目标类别赋值 1 到 16[29]。Elsevier 通过机器学习模型构建了针对 16 个可持续发展目标的训练集，为每个可持续发展目标生成相关主题词及其 TF-IDF 值。TF-IDF 值越高表示该词与这个可持续发展目标关系越密切。为了保证可读性，并未对主题词进行词干提取或词形还原，因此某些主题词可能会重复，如“recycle”“recycling”“industry”“industrial”等。机器学习模型通过补充和完善可持续发展目标主题词，大幅度提高了捕获文章的数量，同时保持精确度在 80% 以上。Elsevier 已将可持续发展目标分类加载到其数据分析平台 Scival 中，并更新了 16 项可持续发展目标出版物。研究人员和机构可通过 Scival 平台跟踪可持续发展目标取得的成就与进展。

Clarivate 对联合国可持续发展目标出版物的获取，主要借助旗下 Citation Topics 的分类体系实现。Clarivate 旗下的 Incites 平台包含多种学科分类法，Citation Topics 是其中之一。Citation Topics 是基于论文层面的文献主题分类法，它打破了文献所在期刊的限制，根据文献之间的直接引用关系，通过莱顿算法[30]对文献进行聚类，并对每个聚类分配一个引文主题，聚类中所有的文献都属于该主题，每篇文献只被分配到一个主题。Citation Topics 包括宏观主题、中观主题和微观主题三层结构，目前宏观主题 10 个、中观主题 326 个、微观主题 2449 个，每个主题由数字标号和主题名称组成。该分类方法首次于 2020 年 12 月在 Incites 平台上推出，并于 2023 年 4 月和 2024 年 4 月进行了两次更新[31]。Clarivate 将微观主题与联合国可持续发展目标进行比较，构建二者之间的映射关系。对映射到每个可持续发展目标的微观主题，其下包含的所有文献都将被划分到相应的可持续发展目标主题中。在 Incites 平台中，每篇可持续发展目标论文会涉及多个引文主题，同样，一些引文主题也会出现在多个可持续发展目标论文中。Incites Citation Topics 中有 196 个微观主题，有的微观主题并非只属于一个可持续发展目标论文，可能会分布在 2 至 3 个可持续发展目标论文中，因此一些文献主题会同时归属于多个可持续发展目标论文。

由此可见，Elsevier 和 Clarivate 对可持续发展目标论文主题的分类存在不同的判断规则。Elsevier 检索系统是基于文献内容，并根据制定的可持续发展目标主题词在论文题名、关键词

和摘要中查找，因此只有包含可持续发展目标主题词的论文才被认为是可持续发展目标论文。Clarivate 检索系统是基于文献之间的引用关系，通过构建微观主题与可持续发展目标的映射以及论文所属的微观主题来判定是否属于可持续发展目标论文。

（2）检索结果存在分类差异

在 Scival 平台上检索到的可持续发展目标论文中，有 375 篇无法在 Incites 平台上检索到，但这些论文是被 Web of Science 核心合集收录的。研究发现，在 2437 个微观主题中，仅有 1898 个微观主题被 Clarivate 映射到可持续发展目标，而这 375 篇论文所属的微观主题都不在这个映射范围内，所以这些论文在 Incites 中就不属于可持续发展目标论文。

在 Incites 平台上检索到的可持续发展目标论文中，有 3190 篇在 Scival 平台上不属于可持续发展目标论文。推测原因可能是两个平台对可持续发展目标的判定标准不同。这些论文中可能并不包括 Elsevier 提供的可持续发展目标主题词，但这些论文的参考文献或施引文献中可能含有大量的可持续发展目标论文。这部分论文涉及的数量较多，且对参考文献和施引文献的查找和分析也有一定难度。

（3）其他

Elsevier 的主题词和 Clarivate 的微观主题在不同的可持续发展目标文献中有重合现象，如"CO_2 reduction"（CO_2 减排）分别存在于 Incites SDG7 和 Scival SDG13 中，"lithium-sulfur batteries"（锂硫电池）和"lithium ion batteries"（锂离子电池）分别存在于 Incites SDG11 和 Scival SDG7 中。由此可见，同一篇论文在 Elsevier 和 Clarivate 中会被划分到不同的可持续发展目标大类。

此外，同一论文可能包含多个可持续发展目标的主题词，这可能导致分类归属偏差的现象。而且，同一论文的研究内容和研究目的可能涉及不同的可持续发展目标，从而属于不同的可持续发展目标大类。

4 存在的问题与改进建议

4.1 存在的问题

到目前为止，关于改进可持续发展目标科学出版物获取方法的讨论尚未达成共识，在理论层面还存在一些模糊认识。一是对可持续发展目标的主题和概念解释不清晰。可持续发展目标是一个多重定义和多重变量相互作用的复杂系统，而且"可持续性"也很难定义，因为它在不同学科中有不同的解释[32]。二是对可持续发展目标研究有贡献的出版物不易界定，包括：出版物如何表述才能确定其研究主题被认为与可持续发展目标有关；可持续发展目标出版物是否有必要对可持续发展目标做出具体的或直接的贡献；产生间接贡献的出版物是否也应该计算在内。三是不同国家和地区的文化传统、语言习惯、经济发展、行业特征及研究侧重点存在结构性差异，但现有的一刀切式搜索查询对这种差异缺乏敏感性，没有在广泛的可持续发展目标框架内充分关注这些细微差别。

在技术层面，无论是布尔逻辑检索还是机器学习算法都有其局限性。基于关键词的布尔逻辑方法搜索词的选择、组合方式及查询结构，都会对检索结果产生影响。来自可持续发展目标官方文件中的关键词虽然质量很高，但往往不全面，不能保证涵盖可持续发展目标相关论述的全部。

选定的关键词是否与可持续发展目标的主题相关，能否解释出版物对可持续发展目标有贡献，以及如何将可持续发展目标的主题概念转换为搜索查询结果，从而准确识别出相关出版物，这些都是需要面对的挑战。一般情况下，监督机器学习模型通常在相对较小且同质的语料库训练，难以处理样本以外的情况。而且，在具有不同参数的不同数据集上训练的模型几乎不可能集成。尽管无监督机器学习有很好的前景，但它目前只应用于相对较小的文本语料库。除此之外，机器学习方法还带来了模型或聚类的可解释性问题[32]。

4.2 改进建议

由此可见，关键词结合机器学习的检索策略（如 Scival 平台）与引文结合内容映射的检索策略（如 Incites 平台），在开展联合国可持续发展目标出版物检索时都有一定的缺陷。针对这两种检索策略存在的局限性，本文尝试从技术方法层面提出如下改进建议。

（1）针对关键词结合机器学习检索策略的改进建议

第一，提供可持续发展目标相关术语并限定研究主题领域，从而减少主题词的多义性。通过将主题词限定在不同的学科或主题内缩小检索范围，将不太相关的文献排除在外。Bordignon[33]就提出建议，在查询 Elsevier 的同时，限制论文在 Scopus 所属学科。

第二，通过正则表达式，统一关键词在文本中出现的多种不同书写形式。例如，表 3 中“dye sensitized solar cell”的不同变体“dye-sensitized solar cells (dscs)”中还存在错误的书写形式，括号中的内容应是“dsscs”，因此可通过构建正则表达式，将其所有书写形式（包括复数、缩写词、词之间的“-”和词尾的“.”）都涵盖其中。正则表达式如下：

“dye(|-)sensiti[a-z]{1,5}(|-)solar(|-)cell(s|)(\.|)(|)($|\(.*?\))”

第三，对检索到的可持续发展文献，需进一步核查其元数据，避免非目标文献的混入。

（2）针对引文结合内容映射的检索策略的改进建议

在 Incites 平台中，有一些未映射或没有被引用过的文献中可能包含可持续发展相关文献，它们在检索时会被忽略。这些文献可通过关键词布尔逻辑查询方法来弥补。

Incites 主要根据文献引用和聚类来识别可持续发展目标文献，缺少对文献语义的分析。因此，可利用 SAO 语义分析法识别并分析句子中的主语、谓语和宾语等要素及其语义关系，进而获知文本的主要信息和关联关系。SAO 语义分析法常用于专利特征分析，也可用于从论文标题、摘要中提取核心内容、特征及语境信息，来判断和识别论文的可持续发展目标类别归属。

对于研究内容和研究目的涉及不同可持续发展目标的论文，可通过引文语境分析法来判断。在 Web of Science 数据库中，可以定位目标文献在所有施引文献中所处的位置，再根据引文语境分析目标文献的价值。该方法也可用于可持续发展目标论文的识别。例如，在文本中先标注出可持续发展目标相关主题词，再根据主题词在文本中的位置（研究背景、研究目的、研究方法、研究结论等）进行分类，最后根据主题词的应用场景来判断相关论文的可持续发展目标类别归属。

5 结 论

构建联合国可持续发展目标出版物的检索策略是一个较为复杂的过程。虽然构建科学出版

物与可持续发展目标的映射关系相对容易，但评估检索策略是否有效却很困难。目前，已有检索策略都存在一定程度的缺陷，而学术界还没有建立统一的检索策略标准。相较于传统的学科与文献分类体系，联合国可持续发展目标出版物分类是一种未经作者和编辑参与的分类体系，很难建立论文出版物与检索文献之间的精准映射关系。因此，需要充分认知上文所述的不同平台分类差异，才可能最大限度地弥补这些缺陷。图书馆应积极开展可持续发展目标出版物检索策略的优化研究，充分挖掘与利用图书馆的数字资源，为教学科研提供面向可持续发展目标的文献资源支撑。

目前，国际数据库检索平台 Elsevier 和 Clarivate 采用不同方法构建了科学出版物与可持续发展目标之间的映射关系。本文在排除数据源收录差异的基础上，分别在这两个平台中检索可持续发展目标论文，比较分析两者存在较大差异的原因并提出改进建议。

实际上，除上述 Scival 和 Incites 等付费平台之外，一些开源数据平台（如 Dimensions）也开展联合国可持续发展目标相关出版物的检索服务。如果对这些平台的可持续发展目标相关出版物开展检索研究，并将其检索结果与上述两个平台的检索结果比对，可能会进一步揭示不同检索策略之间的差异，也可以将其检索结果作为可持续发展目标出版物的有益补充，使检索结果更加全面。

【参考文献】

［1］2030年可持续发展议程联合国可持续发展目标和指标体系解析［EB/OL］.［2023-12-26］. https://sdg.js.org/.

［2］Take a holistic view of the sustainable development goals［EB/OL］.［2025-02-20］. https://2017-2020.usaid.gov/GlobalGoals.

［3］Key messages - Australia［EB/OL］.［2025-02-20］. https://sustainabledevelopment.un.org/memberstates/australia.

［4］中方发布《中国落实2030年可持续发展议程国别方案》［EB/OL］.［2023-11-21］. https://www.gov.cn/xinwen/2016-10/13/content_5118514.htm.

［5］可持续发展大数据国际研究中心［EB/OL］.［2023-11-21］. http://www.cbas.ac.cn/zxjs/zxgk/.

［6］Browse courses by SDG［EB/OL］.［2024-11-15］. https://www.auckland.ac.nz/en/about-us/about-the-university/the-university/sustainability-and-environment/studying-sustainability/browse-courses-by-sdg.html.

［7］SDG全球暑校精彩预告 | 气候治理与可持续发展［EB/OL］.［2023-11-17］. https://mp.weixin.qq.com/s?__biz=MzA3NzUwOTcwNw==&mid=2649189161&idx=1&sn=6e0ba0033810654d783a67b09949dc98&chksm=87432e0db034a71bf85074a747243a765a1d0c41f3fe29b15a5aef9fa3e9ab09eeeb46cb4657&scene=27.

［8］2020复旦大学SDGs行动报告［EB/OL］.［2023-11-16］.https://www.op.fudan.edu.cn/_upload/article/files/09/d5/b70b34524eee925ec90526cb4c30/c1119570-461c-4ec0-9eef-61006f1721d3.pdf.

［9］Armitage C S, Lorenz M, Mikki S. Mapping scholarly publications related to the Sustainable Development Goals: Do independent bibliometric approaches get the same results?［J］. Quantitative Science Studies, 2020, 1(3): 1092-1108.

［10］Bordignon F. Dataset of search queries to map scientific publications to the UN sustainable development goals［J］. Data Brief, 2021(34): 1-12.

［11］Wang W W, Kang W H, Mu J W. Mapping research to the Sustainable Development Goals(SDGs)［EB/OL］.［2023-12-26］. https://www.researchgate.net/publication/368674683_Mapping_research_to_the_Sustainable_Development_Goals_SDGs.

[12] The University of Auckland SDG keywords mapping [EB/OL]. [2023-12-26].https://www.sdgmapping.auckland.ac.nz/.

[13] Universities and the SDGs. [EB/OL]. [2023-12-26]. https://ap-unsdsn.org/regional-initiatives/universities-sdgs/.

[14] Körfgen A, Förster K, Glatz I, et al. It's a hit! Mapping Austrian research contributions to the Sustainable Development Goals [J]. Sustainability/Molecular Diversity Preservation International (MDPI), 2018, 10(9): 1-13.

[15] Confraria H, Noyons E, Ciarli T. Countries' research priorities in relation to the Sustainable Development Goals [C]. 18th International Conference of the International Society for Scientometrics and Informetrics. Leuven, Belgium. 2021: 281-292.

[16] South African SDG Hub [EB/OL]. [2023-12-29]. https://sasdghub.up.ac.za/home/.

[17] Wastl J, Porter S, Draux H, et al. Contextualizing sustainable development research. [EB/OL]. [2024-01-05].https://digitalscience.figshare.com/articles/report/Contextualizing_Sustainable_Development_Research/12200081/2.

[18] Navigating the structure of research on Sustainable Development Goals [EB/OL]. [2024-01-05]. https://clarivate.com/lp/navigating-the-structure-of-research-on-sustainable-development-goals-2/.

[19] LaFleur M. Art is long, life is short: an SDG classification system for DESA publications [EB/OL]. [2023-12-26]. https://www.un.org/esa/desa/papers/2019/wp159_2019.pdf.

[20] Hajikhani A, Suominen A. Mapping the Sustainable Development Goals (SDG) in science, technology and innovation: application of machine learning in SDG-oriented artefact detection [J]. Scientometrics, 2022, 127: 6661-6693.

[21] Zhang R, Vignes M, Steiner U, et al. Matching research publications to the United Nations' Sustainable Development Goals by multi-label-learning with hierarchical categories [C]. IEEE 7th International Conference on Data Science and Advanced Analytics. Sydney, NSW, Australia. 2020: 516-525.

[22] Sustainable Development Goals classification [EB/OL]. [2023-12-29]. https://www.digital-science.com/resource/sustainable-development-goals-classification/.

[23] Fane B, Draux H, Wastl J. Using Digital Science's Dimensions Database to track research with the UN Sustainable Development Goals [C]. 26th International Conference on Science, Technology and Innovation Indicators. Granada, Spain. 2022: 1-8.

[24] Schmidt F, Vanderfeesten M. Evaluation on accuracy of mapping science to the United Nations' Sustainable Development Goals (SDGs) of the Aurora SDG queries [EB/OL]. [2023-12-29]. https://zenodo.org/records/4964606.

[25] Vanderfeesten M, Otten R, Spielberg E. Search queries for "Mapping Research Output to the Sustainable Development Goals (SDGs)" v5.0.2 [EB/OL]. [2024-12-29]. https://doi.org/10.5281/zenodo.4883250.

[26] Maurice V, Robert J, Lennart K. AI for mapping multi-lingual academic papers to the United Nations' Sustainable Development Goals (SDGs) [EB/OL]. [2024-12-29]. https://zenodo.org/records/6487606.

[27] Maxime R, Yury K, Alexandre B, et al. Improving the Scopus and Aurora Queries to identify research that supports the United Nations Sustainable Development Goals (SDGs)2021 [EB/OL]. [2023-12-29]. https://elsevier.digitalcommonsdata.com/datasets/9sxdykm8s4/4.

[28] Kashnitsky Y, Roberge G, Mu J, et al. Evaluating approaches to identifying research supporting the united Nations Sustainable Development Goals [EB/OL]. [2025-02-20]. https://arxiv.org/abs/2209.07285.

[29] SDG research mapping initiative [EB/OL]. [2023-11-16]. https://www.elsevier.com/about/sustainability/sdg-research-mapping-initiative.

[30] Traag V A, Waltman L, Van Eck N J. From louvain to leiden: guaranteeing well-connected communities [J]. Scientific Reports, 2019(9): 1-12.

［31］Citation topics［EB/OL］.［2024-11-19］. https://incites.zendesk.com/hc/en-gb/articles/22514077746961-Citation-Topics.

［32］Tay A. Mapping UN Sustainable Development Goals (SDG) to publications［EB/OL］.［2023-12-22］. https://library.smu.edu.sg/topics-insights/mapping-un-sustainable-development-goals-sdg-publications.

［33］Bordignon F. Search queries to map scientific publications to the UN Sustainable Development Goals［EB/OL］.［2023-12-22］. https://data.mendeley.com/datasets/xrx7ddbbb4/1.

Study on Retrieval Strategies for the United Nations Sustainable Development Goals Publications

Yu Xi

(The Library of Tianjin Normal University, Tianjin 300387, China)

Abstract: ［**Purpose/Significance**］ The United Nations Sustainable Development Goals (SDGs) publications serve as the literature foundation for measuring the contribution of countries, regions, and institutions to research related to SDGs. Optimizing publication retrieval strategies can provide methodological references for related work. ［**Method/Process**］ Based on the Elsevier and Clarivate retrieval system platforms, sustainable development goal related papers were selected as the retrieval objects. The differences in retrieval results between the two databases and the reasons for the differences were compared and analyzed, and improvement suggestions were proposed for the advantages and disadvantages of the two query methods. ［**Result/Conclusion**］ Compared with traditional disciplinary and literature classification systems, the United Nations Sustainable Development Goals (SDGs) publications are a classification method that does not involve authors and editors. Currently, it is difficult to establish an accurate mapping relationship between publications and retrieved literature. Therefore, it is necessary to fully recognize the classification differences of different retrieval system platforms, optimize the retrieval strategy of SDG publications, fully tap and utilize the digital resources of libraries, and provide literature resource support for teaching and research that is oriented towards sustainable development goals.

Keywords: Sustainable Development Goals (SDGs) publications; Search strategy; Elsevier; Clarivate

（本文责编：任全娥）